Building Pixel Classifiers using the Interactive Teacher/Learner (ITL) System*

Justus H. Piater and Adriana Mincheva
Computer Science Department
University of Massachusetts
Amherst, MA 01003
{piater|mincheva}@cs.umass.edu

Abstract

We present an interactive paradigm for the construction of pixel classifiers. The user selects training pixels incrementally, based on real-time feedback from the classifier running in the background. Experiments show that this facilitates the construction of very small yet accurate decision tree classifiers. The framework is extensible in many ways. For example, we show how classifiers can be composed hierarchically and trained to find seagulls in aerial images.

1. Introduction to the ITL concept

Pixel classifiers are an important component of many vision systems, e.g. for texture-based segmentation, obstacle detection, or geoscience. Despite these abundant applications, the construction of high-performance pixel classifiers usually involves substantial cost in terms of human effort. A major objective of this work is to simplify the selection of training examples during the process of classifier construction. In most applications today, large amounts of training data are prepared a priori. This is expensive since ground truth information usually must be provided by hand. Moreover, the contribution of individual training pixels to the final classification is not known. Evidently, if one could choose the instances appropriately, very similar classification performance could be achieved with much fewer training examples.

We propose to supply training instances incrementally: If one knows where the classifier currently makes mistakes, one can generate an *informative* training instance by providing a correct label for a currently misclassified pixel. We employ an interactive Teacher-Learner paradigm (Figure 1) for incremental selection of training pixels, which we call ITL. The Teacher is a human domain expert who operates a graphical user interface. He can select images for training and, for any image, select and label small clusters of pixels. The Learner is a computer program that operates through a well-defined communication interface with the Teacher. The Learner can receive images and training instances, can quickly produce and update a classifier, and generates labels for the pixels of the current training image. Each time the user provides a new instance, the Learner rapidly revises its classifier as necessary, and then recomputes the class labels for all pixels of the image. This lets the user see the misclassified pixels with almost no delay. He can immediately respond by providing correct labels for one or more of them, which are passed as new training examples to the classifier.

Our system employs Utgoff's Incremental Tree Inducer ITI [3], a state-of-the-art decision tree algorithm. One of its unique features is that it can incorporate training instances serially without needing to rebuild the tree repeatedly. Training examples, once provided, are never forgotten and will always be classified correctly, unless indistinguishable pixels with contradictory labels are provided. Moreover, it produces the same tree for the same accumulated set of training instances, regardless of the order in which they are received. It achieves a very quick feedback loop, consisting of receiving a new training instance, updating the classifier, and reclassifying the image. This allows ITI to function in close to real time, ensuring the effectiveness of the human in the loop.

To maximize the utility to the user, pixels near the location of the latest training pixel are (re)classified first and displayed by our graphic user interface. The user can select new training pixels at any time, without waiting for the clas-

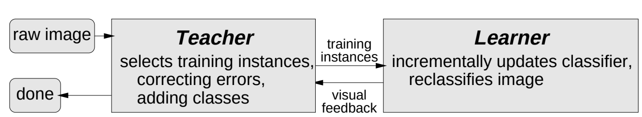

Figure 1. The ITL framework: interactive, incremental classifier construction.

*This research was supported in part by ARPA (via ARL) under contracts DAAL02-91-K-0047 and DACA76-97-K-0005, and by NSF under grants IRI-9222766 and EIA-9726401.

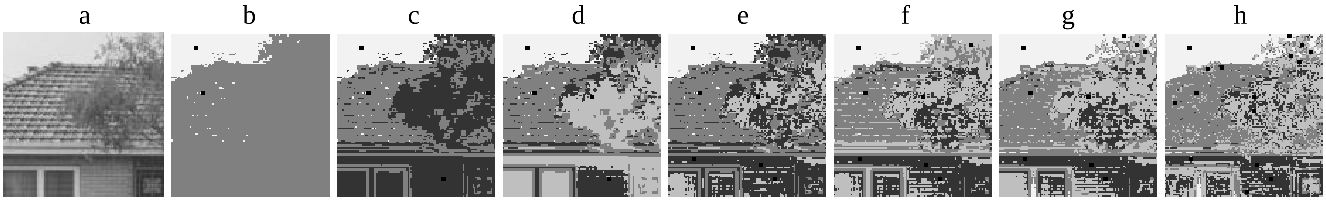

Figure 2. An example training session: (a) gray-scale version of the original color image; (b)–(d): results after adding one set of training instances for each class; (e)–(h): snapshots during some refining. Training pixels appear as tiny black squares. sky, roof, foliage, brick.

sification process to complete. This facilitates very rapid training even on large images.

2. Examples

Figure 2 shows snapshots of an interactive session [2]. The goal is to train a classifier to label pixels as sky, roof, brick, or foliage. Six features are used, which are the raw RGB values of a pixel, and the variances of each in a 3×3 window centered around that pixel. Each mouse click of the teacher produces a training instance of a specific class that is used to update the learner's decision tree. Immediately, the Learner begins to reclassify the image using the updated tree. Figure 2 shows various intermediate stages of training.

On a large, more realistic terrain classification task, we quantified the performance of classifiers produced using the ITL framework. Here, interactively constructed classifiers achieved more than 97% of the accuracy of a conventionally trained ITI classifier [2]. The trees produced by ITL were many times smaller and tested less than half as many features, which promises good generalization properties. In our experiments to date, training a classifier typically requires less than 10 mouse clicks per class, and takes only a few minutes.

Hierarchical classification [1] allows the user to train a classifier to recognize structured objects. For example, seagulls are characterized by adjacent clusters of white and dark pixels, corresponding to head and back (Figure 3a, b). At the base level of a hierarchy of classifiers, a decision tree is constructed as described above. For example, in Figure 3c a classifier was trained to recognize the white head and dark back of a seagull. Note the many false positives in both classes. At the next level, another classifier is trained whose inputs are features computed from the labels generated by the base-level classifier. Very general such features can be defined to express spatial relationships between clusters of class labels [1]. In Figure 3d, gulls are recognized – without errors – by the adjacency of appropriately sized clusters of "head" and "back" labels. This hierarchy of classifiers was trained simply by clicking on positive and negative examples of gulls and their parts. In principle, the hierarchy can be extended to any number of classifiers.

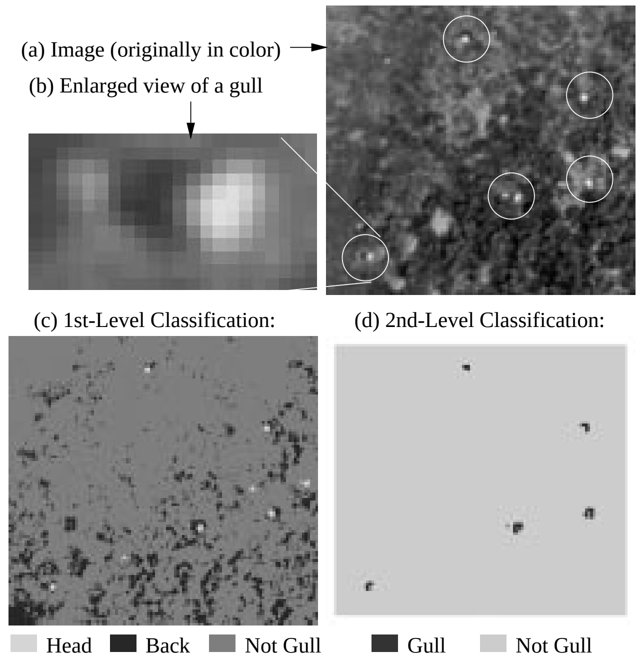

Figure 3. Finding seagulls using hierarchical classification.

Current work includes a large-scale environmental monitoring project. To aid the user in the specification of ground truth examples, we are experimenting with animated 3-D fly-through terrain visualization.

References

[1] J. H. Piater, E. M. Riseman, and P. E. Utgoff. Interactive training of pixel classifiers opens new possibilities. In *ISPRS ComIII Symposium on Object Recognition and Scene Classification from Multispectral and Multisensor Pixels*, International Archives of Photogrammetry and Remote Sensing, Columbus, OH, July 1998.

[2] J. H. Piater, E. M. Riseman, and P. E. Utgoff. Interactively training pixel classifiers. *International Journal of Pattern Recognition and Artificial Intelligence*, to appear.

[3] P. E. Utgoff, N. C. Berkman, and J. A. Clouse. Decision tree induction based on efficient tree restructuring. *Machine Learning*, 29(1):5–44, October 1997.